

# Rate-Distortion Dimension of Stochastic Processes

Farideh Ebrahim Rezagah\*, Shirin Jalali<sup>†</sup>, Elza Erkip\*, H. Vincent Poor<sup>‡</sup>

\* NYU Tandon School of Engineering, {farideh,elza}@nyu.edu

<sup>†</sup> Nokia - Bell Labs, Shirin.Jalali@alcatel-lucent.com

<sup>‡</sup> Electrical Engineering Department, Princeton University, poor@princeton.edu

**Abstract**—The rate-distortion dimension (RDD) of an analog stationary process is studied as a measure of complexity that captures the amount of information contained in the process. It is shown that the RDD of a process, defined as two times the asymptotic ratio of its rate-distortion function  $R(D)$  to  $\log \frac{1}{D}$  as the distortion  $D$  approaches zero, is equal to its information dimension (ID). This generalizes an earlier result by Kawabata and Dembo and provides an operational approach to evaluate the ID of a process, which previously was shown to be closely related to the effective dimension of the underlying process and also to the fundamental limits of compressed sensing. The relation between RDD and ID is illustrated for a piecewise constant process.

**Index Terms**—Rate-Distortion Dimension, Information Dimension, Compressed Sensing

## I. INTRODUCTION

For discrete-alphabet signals, the Shannon entropy function  $H(X)$  and the entropy rate  $\bar{H}(\mathbf{X}) = \lim_{n \rightarrow \infty} H(X_n | X^{n-1})$  measure the complexity of a random variable  $X$  and a stationary stochastic process  $\mathbf{X} = \{X_i\}$ , respectively. Both of these measures are closely connected to the minimum number of bits per symbol required for representing stochastic sources [2] and can also be thought of as measures of signal structure. However, when we shift from discrete alphabet to continuous alphabet, both the entropy and the entropy rate become infinite. Instead, for analog signals, the notion of information dimension (ID) introduced by Rényi [3] provides a framework that can be used to quantify signal structure.

To illustrate what is meant for an analog process to be structured, consider a stationary memoryless (i.e., independent and identically distributed or i.i.d.) process  $\mathbf{X} = \{X_i\}_{i=0}^{\infty}$  such that  $X_i \sim (1-p)\delta_0 + pf_c$ , where  $f_c$  denotes the probability density function (pdf) of an absolutely continuous distribution and  $\delta_0$  denotes the Dirac measure with an atom at 0. In other words, for each  $i$ , with probability  $p \in [0, 1]$ ,  $X_i$  is exactly equal to zero; otherwise, it is drawn from  $f_c$ . By the strong law of large numbers, for large values of blocklength  $n$ , with probability approaching one, a block  $X^n$  generated by this source contains around  $n(1-p)$  entries equal to zero, and the rest of the entries are real numbers in the domain of  $f_c$ . To describe  $X^n$  with a certain precision, for zero entries, it suffices to describe their locations. The

number of bits required for this description does not depend on the reconstruction quality. However, for the remaining approximately  $np$  elements of  $X^n$ , it is known from rate-distortion theory that the required number of bits grows with the desired reconstruction quality. This intuitively suggests that  $p$ , which controls the number of non-zero elements in  $X^n$ , is a fundamental quantity related to the complexity and structure of  $X^n$ . This intuition is accurately captured by the ID of this source which can be shown to be equal to  $p$  [3]. In fact,  $\delta_0$  can be changed to any discrete probability distribution with finite entropy and the result will not change since the Rényi ID of a discrete source is 0.

A further significance of the ID as a measure of structure is its relationship to the problem of compressed sensing. Consider the problem of recovering a signal  $X_o^n$  from under-determined measurements  $Y^m = AX_o^n$ , where  $m < n$ . It is known that if the input signal  $X_o^n$  is sparse, or in general “structured”, it can be accurately recovered from the measurements, even if  $m$  is far fewer than  $n$  [4]–[9]. For stationary memoryless processes, under some mild conditions on the distribution, the Rényi ID of the first order marginal distribution of the source characterizes the fundamental limits of compressed sensing, i.e., the minimum number of measurements required for asymptotically almost lossless recovery [10]. The notion of the Rényi ID is extended to stationary processes in [11], where it is proved that there is a direct relationship between the ID of a stationary process and the number of random linear measurements required for its universal recovery.

While the aforementioned results give an operational meaning to the ID of a signal, evaluating the ID of a stationary process is in general difficult. Kawabata and Dembo defined the rate-distortion dimension (RDD) of i.i.d. random variables (or vectors) based on the rate-distortion trade-off in the asymptotically low distortion regime [12]. They proved that for a random variable, its (upper and lower) RDD is equal to its (upper and lower) ID.

The main contribution of this paper is to extend the notion of RDD to analog stationary processes, and to prove that, under some regularity conditions, the RDD of a stationary process is equal to its ID, defined in [11]. This provides an extension of the result of Kawabata and Dembo to stochastic processes, and thereby provides a computationally feasible way of finding the ID of a stochastic process. In order to illustrate this, we compute the RDD of piecewise-constant stochastic processes,

This work is part of a paper under review by the *IEEE Transactions on Information Theory*, available at [1]. This research was supported by the National Science Foundation under Grant CCF-1420575.

which are widely used to model many natural signals such as images. We derive upper and lower bounds on the rate-distortion functions of such signals, and use these bounds to evaluate the RDD and hence, the ID of such processes. Furthermore, our results in [1] suggest that the RDD of a stochastic process is closely related to the fundamental limits of compressed sensing for the process, and hence RDD and ID can be thought of as measures of structure/complexity for arbitrary stationary stochastic processes.

The organization of the paper is as follows. Section II defines and examines the properties of ID and RDD. Section III contains our main result which establishes a connection between the ID and the RDD of stochastic processes. Upper and lower bounds on the rate-distortion region of the piecewise constant source modeled by a first-order Markov process are provided in Section IV to illustrate the relationship between RDD and ID. Section V concludes the paper.

### A. Notation

Capital letters like  $X$  and  $Y$  represent random variables. For  $x \in \mathbb{R}$ ,  $\lceil x \rceil$  ( $\lfloor x \rfloor$ ) represents the smallest (largest) integer larger (smaller) than  $x$ . For  $b \in \mathbb{N}^+$ ,  $[x]_b$  denotes the  $b$ -bit approximation of  $x$ , i.e., for  $x = \lfloor x \rfloor + \sum_{i=1}^{\infty} (x)_i 2^{-i}$ ,  $(x)_i \in \{0, 1\}$ ,  $[x]_b = \lfloor x \rfloor + \sum_{i=1}^b (x)_i 2^{-i}$ . Also, let  $\langle x \rangle_b$  be defined as  $\langle x \rangle_b = \frac{\lfloor bx \rfloor}{b}$ . For  $x^n \in \mathbb{R}^n$ ,  $[x^n]_b$  and  $\langle x^n \rangle_b$  are defined as  $([x_1]_b, \dots, [x_n]_b)$  and  $(\langle x_1 \rangle_b, \dots, \langle x_n \rangle_b)$ , respectively. Throughout the paper, log refers to the logarithm in base 2.

## II. BACKGROUND

In this section, we provide formal definitions of ID and RDD and an overview of the literature.

**Definition 1** (Rényi information dimension [3]). *The Rényi upper and lower IDs of an analog random variable  $X$  are defined as*

$$\bar{d}(X) = \limsup_{b \rightarrow \infty} \frac{H(\langle X \rangle_b)}{\log b},$$

and  $\underline{d}(X) = \liminf_{b \rightarrow \infty} \frac{H(\langle X \rangle_b)}{\log b}$ , respectively. If the two limits coincide, the Rényi ID of  $X$  is defined as  $d(X) = \bar{d}(X) = \underline{d}(X)$ .

Definition 1 can also be applied to analog vectors. For instance, for a random vector  $X^n$ ,  $\bar{d}(X^n) = \limsup_{b \rightarrow \infty} \frac{H(\langle X^n \rangle_b)}{\log b}$ .

While the above definition of the Rényi ID is in terms of the entropy of the  $b$ -level quantized version of  $X$  normalized by the number of bits required for binary representation of it,  $\log b$ , as proved in Proposition 2 of [10], it can equivalently be defined in terms of the entropy of the  $b$ -bit quantized version of  $X$ ,  $[X]_b$ , normalized by  $b$ , i.e.,

$$\bar{d}(X) = \limsup_{b \rightarrow \infty} \frac{H([X]_b)}{b},$$

and  $\underline{d}(X) = \liminf_{b \rightarrow \infty} \frac{H([X]_b)}{b}$ .

The notion of Rényi ID for random variables or vectors was extended in [11] to define the ID of analog stationary processes.

**Definition 2** (ID of a stationary process [11]). *The  $k$ -th order upper and lower IDs of stationary process  $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$  are defined as*

$$\bar{d}_k(\mathbf{X}) = \limsup_{b \rightarrow \infty} \frac{1}{b} H([X_{k+1}]_b | [X^k]_b),$$

and  $\underline{d}_k(\mathbf{X}) = \liminf_{b \rightarrow \infty} \frac{1}{b} H([X_{k+1}]_b | [X^k]_b)$ , respectively. The upper and lower ID of the process  $\mathbf{X}$  are defined as

$$\bar{d}_o(\mathbf{X}) = \lim_{k \rightarrow \infty} \bar{d}_k(\mathbf{X})$$

and  $\underline{d}_o(\mathbf{X}) = \lim_{k \rightarrow \infty} \underline{d}_k(\mathbf{X})$ , respectively, when the limits exist. If  $\bar{d}_o(\mathbf{X}) = \underline{d}_o(\mathbf{X})$ , the ID of process  $\mathbf{X}$ ,  $d_o(\mathbf{X})$ , is defined as  $d_o(\mathbf{X}) = \bar{d}_o(\mathbf{X}) = \underline{d}_o(\mathbf{X})$ .

As proved in [11], both  $\bar{d}_k(\mathbf{X})$  and  $\underline{d}_k(\mathbf{X})$  are both non-negative decreasing sequences in  $k$ . Hence, if they are also bounded, which is the case for instance for bounded sources, their limits as  $k \rightarrow \infty$  also exist.

For a stationary memoryless process  $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$ , this definition coincides with that of Rényi's ID of the first-order marginal distribution of the process  $\mathbf{X}$ . That is  $\bar{d}_o(\mathbf{X}) = \bar{d}(X_1)$  and  $\underline{d}_o(\mathbf{X}) = \underline{d}(X_1)$ . For sources with memory, taking the limit as the memory parameter  $k$  grows to infinity allows  $d_o(\mathbf{X})$  to capture the overall structure that is present in an analog stationary process. It can be proved that  $d_o(\mathbf{X}) \leq 1$ , for all bounded stationary processes, and if the stationary process  $\mathbf{X}$  is structured,  $d_o(\mathbf{X})$  is strictly smaller than one [11].

Under some mild conditions on the distribution, [10] proves that the Rényi ID of the first-order marginal distribution of a stationary memoryless process characterizes the fundamental limits of its compressed sensing. In other words, given a stationary memoryless process  $\mathbf{X}$ , asymptotically, as the blocklength  $n$  grows to infinity, the minimum number of linear projections ( $m$ ) normalized by the blocklength ( $n$ ) that is required for recovering source  $X^n$  is shown to be equal to  $d(X_1)$ . In [11], it is shown that, asymptotically, slightly more than  $n \bar{d}_o(\mathbf{X})$  random linear projections suffice for *universal* recovery of  $X^n$  generated by any stationary process that satisfies some mixing conditions. These results provide an operational interpretation of the ID of a random process.

The rate-distortion function of a stationary source measures the minimum number of bits per source symbol required for achieving a given reconstruction quality. In some cases, as the reconstruction becomes finer, the behavior of the rate-distortion function is connected to the level of structuredness of the source process and also to its ID mentioned earlier. In the rest of this section, we review the known results on these connections.

Consider a metric space  $(\mathbb{R}^k, \rho)$ , and random vector  $X^k$ . The standard rate-distortion function [2] of vector  $X^k$  under distortion measure  $d(x^k, \hat{x}^k) = \rho(x^k, \hat{x}^k)^r$ , where  $r > 0$ , is

defined as

$$R_r(X^k, D) = \inf_{\mathbb{E}[d(X^k, \hat{X}^k)] \leq D} I(X^k; \hat{X}^k).$$

**Definition 3** (Rate-distortion dimension (RDD) of a random vector [12]). *The upper and lower RDDs of  $X^k$  are defined as*

$$\overline{\dim}_R(X^k) = r \limsup_{D \rightarrow 0} \frac{R_r(X^k, D)}{\log \frac{1}{D}},$$

and  $\underline{\dim}_R(X^k) = r \liminf_{D \rightarrow 0} \frac{R_r(X^k, D)}{\log \frac{1}{D}}$ , respectively. If  $\overline{\dim}_R(X^k) = \underline{\dim}_R(X^k)$ , the RDD of  $X^n$  is defined as  $\dim_R(X^k) = r \lim_{D \rightarrow 0} \frac{R_r(X^k, D)}{\log \frac{1}{D}}$ .

The following theorem from [12] establishes the connection between the Rényi ID of a random vector  $X^k$  and its RDD, for a general distribution on  $X^k$ .

**Theorem 1** (Proposition 3.3 in [12]). *Consider the metric space  $(\mathbb{R}^k, \rho)$ , such that there exists  $0 < a_1 \leq a_2 < \infty$  for which  $a_1 \max_{i=1}^k |x_i - \hat{x}_i| \leq \rho(x^k, \hat{x}^k) \leq a_2 \max_{i=1}^k |x_i - \hat{x}_i|$ , for all  $x^k, \hat{x}^k \in \mathbb{R}^k$ . Then, for any distribution of  $X^k$ ,*

$$\overline{\dim}_R(X^k) = \bar{d}(X^k),$$

and  $\underline{\dim}_R(X^k) = \underline{d}(X^k)$ , where  $\overline{\dim}_R(X^k)$ , and  $\underline{\dim}_R(X^k)$  denote the upper and lower RDD of  $X^k$  under fidelity constraint  $d(x^k, \hat{x}^k) = \rho(x^k, \hat{x}^k)^r$ .

### III. EQUIVALENCE OF RDD AND ID FOR ANALOG PROCESSES

This section provides the main result of this paper which extends the notion of RDD to stationary processes and establishes its connection of the ID of the process.

Consider an analog stationary process  $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$ . The rate-distortion function  $R(\mathbf{X}, D)$  of the source  $\mathbf{X}$  under squared error distortion can be characterized as [13], [14]

$$R(\mathbf{X}, D) = \lim_{m \rightarrow \infty} R^{(m)}(\mathbf{X}, D),$$

where

$$R^{(m)}(\mathbf{X}, D) = \inf_{\mathbb{E}[d_m(X^m, \hat{X}^m)] \leq D} \frac{1}{m} I(X^m; \hat{X}^m)$$

and

$$d_m(x^m, \hat{x}^m) = \frac{1}{m} \|x^m - \hat{x}^m\|_2^2. \quad (1)$$

Note that with this distortion metric, we have  $r = 2$  and  $R^{(m)}(\mathbf{X}, D) = \frac{1}{m} R_2(X^m, D)$ . It can also be shown that  $\inf_m R^{(m)}(\mathbf{X}, D) = R(\mathbf{X}, D)$  [14].

**Definition 4** (RDD of a stationary process). *The upper and lower RDDs of a stationary process  $\mathbf{X}$  is defined as*

$$\overline{\dim}_R(\mathbf{X}) = 2 \limsup_{D \rightarrow 0} \frac{R(\mathbf{X}, D)}{\log \frac{1}{D}}$$

and  $\underline{\dim}_R(\mathbf{X}) = 2 \liminf_{D \rightarrow 0} \frac{R(\mathbf{X}, D)}{\log \frac{1}{D}}$ . If  $\overline{\dim}_R(\mathbf{X}) = \underline{\dim}_R(\mathbf{X})$ , then the RDD of  $\mathbf{X}$  is defined as  $\dim_R(\mathbf{X}) = \overline{\dim}_R(\mathbf{X}) = \underline{\dim}_R(\mathbf{X})$ .

The following theorem extends the equivalence of Rényi ID and RDD established in Theorem 1 for i.i.d. random vectors to stationary processes.

**Theorem 2.** *For a stationary process  $\mathbf{X} = \{X_i\}_{i=-\infty}^{\infty}$ , assume that  $\lim_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}}$  exists for all  $m$ . Then,*

$$\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X}).$$

The main ingredients of the proof of Theorem 2 are the following two lemmas.

**Lemma 1.** *For any stationary process  $\mathbf{X}$ , we have*

$$\overline{\dim}_R(\mathbf{X}) \leq \bar{d}_o(\mathbf{X}) \leq \inf_m 2 \left( \limsup_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right).$$

**Lemma 2.** *Assume that  $\lim_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}}$  exists for all  $m$ , and also there exists  $\sigma_{\max}^2 > 0$ , such that  $R^{(m)}(\mathbf{X}, D)$  uniformly converges to  $R(\mathbf{X}, D)$ , for  $D \in (0, \sigma_{\max}^2)$ , as  $m$  grows to infinity. Then,  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X})$ .*

*Proof of Lemma 1:* Given  $k$ , define distance measure  $\rho_k$  such that for  $x^k, \hat{x}^k \in \mathbb{R}^k$ ,  $\rho_k(x^k, \hat{x}^k) \triangleq \sqrt{k} d_k(x^k, \hat{x}^k)$  where  $d_k(\cdot, \cdot)$  is defined in (1). Note that  $(\mathbb{R}^k, \rho_k)$  is a metric space. Furthermore, since  $\max_{i=1}^k |x_i - \hat{x}_i| \leq \rho_k(x^k, \hat{x}^k) \leq \sqrt{k} \max_{i=1}^k |x_i - \hat{x}_i|$ , from Theorem 1,

$$2 \limsup_{D \rightarrow 0} \frac{k R^{(k)}(\mathbf{X}, \frac{D}{k})}{\log \frac{1}{D}} = \bar{d}(X^k).$$

By a change of variable,  $2 \limsup_{D \rightarrow 0} \frac{k R^{(k)}(\mathbf{X}, D)}{\log \frac{1}{D} + \log \frac{1}{k}} = \bar{d}(X^k)$ , or

$$2 \limsup_{D \rightarrow 0} \frac{R^{(k)}(\mathbf{X}, D)}{\log \frac{1}{D}} = \frac{1}{k} \bar{d}(X^k).$$

Taking the limit of both sides as  $k$  grows to infinity, and employing Lemma 2 from [11], which shows that the upper ID of a process  $\mathbf{X}$  can alternatively be represented as

$$\bar{d}_o(\mathbf{X}) = \lim_{k \rightarrow \infty} \frac{1}{k} \left( \limsup_{b \rightarrow \infty} \frac{H([X^k]_b)}{b} \right),$$

yields

$$\lim_{k \rightarrow \infty} \left( 2 \limsup_{D \rightarrow 0} \frac{R^{(k)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) = \lim_{k \rightarrow \infty} \frac{1}{k} \bar{d}(X^k) = \bar{d}_o(\mathbf{X}). \quad (2)$$

Since  $R^{(k)}(\mathbf{X}, D) \geq \inf_m R^{(m)}(\mathbf{X}, D)$ , from (2),

$$\begin{aligned} \bar{d}_o(\mathbf{X}) &\geq \lim_{k \rightarrow \infty} \left( 2 \limsup_{D \rightarrow 0} \frac{\inf_m R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) \\ &\stackrel{(a)}{=} \lim_{k \rightarrow \infty} \left( 2 \limsup_{D \rightarrow 0} \frac{R(\mathbf{X}, D)}{\log \frac{1}{D}} \right) = \overline{\dim}_R(\mathbf{X}), \end{aligned}$$

where (a) follows from the fact that  $R(\mathbf{X}, D) = \inf_m R^{(m)}(\mathbf{X}, D)$  [14]. This proves the lower bound in the desired result.

To prove the upper bound, fix a positive integer  $m \in \mathbb{N}$ . Any integer  $k$  can be written as  $k = sm + r$ , where  $r \in \{0, \dots, m-1\}$ .

1}. Since  $kR^{(k)}(\mathbf{X}, D)$  is a sub-additive sequence [14], and  $k = m + \dots + m + r$ ,  $kR^{(k)}(\mathbf{X}, D) \leq smR^{(m)}(\mathbf{X}, D) + rR^{(r)}(\mathbf{X}, D)$ , it follows that or

$$R^{(k)}(\mathbf{X}, D) \leq \frac{sm}{k} R^{(m)}(\mathbf{X}, D) + \frac{r}{k} R^{(r)}(\mathbf{X}, D). \quad (3)$$

Combining (2) and (3), it follows that

$$\begin{aligned} \bar{d}_o(\mathbf{X}) &\leq 2 \lim_{k \rightarrow \infty} \left( \limsup_{D \rightarrow 0} \frac{sm}{k} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) \\ &\quad + 2 \lim_{k \rightarrow \infty} \left( \limsup_{D \rightarrow 0} \frac{r}{k} \frac{R^{(r)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) \\ &= 2 \lim_{k \rightarrow \infty} \left( \frac{sm}{k} \right) \left( \limsup_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) \\ &\quad + 2 \lim_{k \rightarrow \infty} \left( \frac{r}{k} \right) \left( \limsup_{D \rightarrow 0} \frac{R^{(r)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right) \\ &= 2 \left( \limsup_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right). \end{aligned} \quad (4)$$

Since  $m$  is selected arbitrarily, we can take the infimum of the right hand side of (4) and derive the desired result. ■

*Proof of Lemma 2:* By the lemma's assumption,  $\underline{\dim}_R(\mathbf{X}) = \dim_R(\mathbf{X})$ ; therefore, from Lemma 1,

$$\dim_R(\mathbf{X}) \leq \bar{d}_o(\mathbf{X}) \leq 2 \left( \lim_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \right), \quad (5)$$

for all  $m$ . Given the uniform convergence assumption, for any  $\epsilon > 0$ , there exists  $m_\epsilon \in \mathbb{N}$ , such that for all  $m > m_\epsilon$ ,

$$\left| \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} - \frac{R(\mathbf{X}, D)}{\log \frac{1}{D}} \right| < \epsilon, \quad (6)$$

for all  $D \in (0, \sigma_{\max}^2)$ .

On the other hand, for any  $\epsilon' > 0$  and  $m$ , there exists  $\delta_{\epsilon', m} > 0$ , such that for all  $D \in (0, \delta_{\epsilon', m})$ ,

$$\lim_{D \rightarrow 0} \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} \leq \frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}} + \epsilon'. \quad (7)$$

Also, for any  $\epsilon'' > 0$ , there exists  $\delta_{\epsilon''} > 0$ , such that for all  $D \in (0, \delta_{\epsilon''})$ ,

$$\frac{R(\mathbf{X}, D)}{\log \frac{1}{D}} \leq \frac{1}{2} (\dim_R(\mathbf{X}) + \epsilon''). \quad (8)$$

Therefore, for any  $\epsilon, \epsilon'$  and  $\epsilon''$ , choosing  $m > m_\epsilon$ , and  $D \in (0, \min(\delta_{\epsilon', m}, \delta_{\epsilon''}))$ , and combining (6), (7) and (8) yields

$$\bar{d}_o(\mathbf{X}) \leq \dim_R(\mathbf{X}) + \epsilon + \epsilon' + \epsilon''. \quad (9)$$

Since  $\epsilon, \epsilon'$  and  $\epsilon''$  are selected arbitrarily, combining (5) and (9) proves that  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X})$ . ■

*Proof of Theorem 2:* It is shown in [15] that for any stationary process  $\mathbf{X}$

$$|R^{(m)}(\mathbf{X}, D) - R(\mathbf{X}, D)| \leq \frac{1}{m} I(X^m; X_{-\infty}^0). \quad (10)$$

Note that while some of the results in [15] hold only for sources that are either absolutely continuous or discrete, as

shown in [1], this bound holds for sources with general distributions. Since the right hand side of (10) does not depend on  $D$ , it shows that  $R^{(m)}(\mathbf{X}, D)$  uniformly converges to  $R(\mathbf{X}, D)$  for all  $D > 0$ . On the other hand, for any  $0 < \sigma_{\max} < 1$ , and any  $D \in (0, \sigma_{\max}^2)$ ,  $0 < 1/\log \frac{1}{D} < 1/\log \frac{1}{\sigma_{\max}^2}$ . Therefore,  $\frac{R^{(m)}(\mathbf{X}, D)}{\log \frac{1}{D}}$  uniformly converges to  $\frac{R(\mathbf{X}, D)}{\log \frac{1}{D}}$ , for  $D \in (0, \sigma_{\max}^2)$ , and by Lemma 2,  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X})$ . ■

For an i.i.d. source  $\mathbf{X}$ , under some mild conditions,  $d_o(\mathbf{X})$  characterizes the fundamental limits of compressed sensing [10]. In other words, asymptotically, almost lossless recovery of  $X^n$  generated by the source  $\mathbf{X}$  from measurements  $Y^m = AX^n$  is feasible, if and only if the normalized number of measurements  $(m/n)$  is larger than  $d_o(\mathbf{X})$ . If the rate-distortion function of the source satisfies the condition of Theorem 2, then  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X})$ , which implies that the RDD of an i.i.d. process can also be used to characterize its compressed sensing fundamental limits. On the other hand, compression-based compressed sensing of stochastic processes is studied in [1]. It is shown in [1] that there exists a compression-based recovery algorithm that achieves almost lossless recovery by using slightly more than  $n \underline{\dim}_R(\mathbf{X})$  random linear measurements. This implies that  $\underline{\dim}_R(\mathbf{X})$  is achievable for general sources. (Note that, by Lemma 1, in general  $\underline{\dim}_R(\mathbf{X}) \leq \bar{d}_o(\mathbf{X})$ .)

**Remark 1.** Theorem 2, by proving the equivalence of ID and RDD, provides a potentially easier path to computing the ID of stochastic processes. Note that while to directly compute the ID of a process one needs to take the limit over the quantized approximations and then over the memory length, to be able to calculate the RDD of a process, the exact characterization of the rate-distortion function is not required. In fact, it is easy to see that it would be enough to have upper and lower bounds on the rate-distortion function of the source,  $R(\mathbf{X}, D)$ , that are within a reasonable gap. More precisely, as long as the gap between the bounds grows as  $o(\log \frac{1}{D})$ , they can be used to evaluate the RDD. Moreover, since the RDD depends only on the low-distortion behavior of the rate-distortion function, studying its asymptotic small distortion performance is sufficient for computing the RDD, and as by Theorem 2, ID of a source, without knowing the rate-distortion function explicitly. For instance, [16] studies the asymptotic behavior of the rate-distortion function of some stochastic sources and employs those results to evaluate the RDD of some i.i.d. processes.

The next section illustrates computation of RDD and its relation to ID for a piecewise constant process.

#### IV. RDD OF A PIECEWISE-CONSTANT PROCESS

In general, deriving the rate-distortion function of sources with memory is very challenging. For instance, even for the binary symmetric Markov chain, the rate-distortion function is not known, except in a low-distortion region [17], and we have to resort to upper and lower bounds [18], [19].

In this section we consider a piecewise constant signal



modeled by a first order Markov process  $\mathbf{X} = \{X_i\}_{i=1}^\infty$ , such that conditioned on  $X_{i-1} = x_{i-1}$ ,  $X_i$  is distributed according to  $(1-p)\delta_{x_{i-1}} + pf_c$ , where  $f_c$  denotes the pdf of an absolutely continuous distribution with bounded support, defined over an interval  $(l, u)$ . In other words, at each time  $i$ , the process either makes a jump and takes a value drawn from distribution  $f_c$ , or it stays at  $X_{i-1}$ . The decision is made based on the outcome of an i.i.d. Bern( $p$ ) random process independent of all past values of  $\mathbf{X}$ . While the output of this source is not sparse, it is clearly a structured process. The following theorem provides upper and lower bounds on  $R(\mathbf{X}, D)$  of the piecewise-constant source. While there is a gap between the bounds on  $R(\mathbf{X}, D)$ , since the gap does not depend on  $D$ , as shown in the following corollary, they can be used to evaluate RDD of the source exactly.

**Theorem 3.** *Consider a first-order stationary Markov process  $\mathbf{X} = \{X_i\}_{i=0}^\infty$ , such that conditioned on  $X_{i-1} = x_{i-1}$ ,  $X_i$  is distributed according to  $(1-p)\delta_{x_{i-1}} + pf_c$ , where  $f_c$  denotes the pdf of an absolutely continuous distribution with bounded support,  $(l, u)$ . If  $d_{\max} \triangleq \sup_{x, \hat{x} \in (l, u)} d(x, \hat{x}) < \infty$ , then*

$$pR_{f_c}(D) \leq R(\mathbf{X}, D) \leq H(p) + pR_{f_c}(D),$$

where  $R_{f_c}(D)$  and  $H(p)$  denote the rate distortion function of an i.i.d. process distributed according to pdf  $f_c$ , and the binary entropy function  $(-p \log_2 p - (1-p) \log_2 (1-p))$ , respectively.

*Proof:* To prove the upper bound (achievability), we consider a code that describes the positions of the jumps losslessly at rate  $H(p)$ . Since the source is piecewise constant, after describing the positions of the jumps, the encoder removes the repeated values and applies a lossy compression code of blocklength close to  $np$ . Therefore, to describe the values at distortion  $D$  the encoder roughly needs to spend  $npR_{f_c}(D)$  bits. For the lower bound (converse), we consider a genie-aided decoder that has access to the positions of the jumps. Then intuitively, to describe the values at distortion  $D$ , it still needs a rate of at least  $pR_{f_c}(D)$ . The proof presented in [1] makes these steps formal by properly analyzing the reduced block length which is a random number. ■

**Corollary 1.** *For the piecewise constant source in Theorem 3, we have  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X}) = p$ . In other words, the RDD is equal to  $p$  which is in turn equal to the ID of this source.*

*Proof:* Given the bound on the rate-distortion process derived in Theorem 3, it is easy to directly derive the RDD of such a source. More precisely, given the upper bound, it follows that

$$\overline{\dim}_R(\mathbf{X}) = 2 \limsup_{D \rightarrow 0} \frac{R(\mathbf{X}, D)}{\log \frac{1}{D}} = p \left( \limsup_{D \rightarrow 0} \frac{R_{f_c}(D)}{\log \frac{1}{D}} \right) = p,$$

where the last step follows from [3] and [12]. Similarly, given the lower bound, we have  $\underline{\dim}_R(\mathbf{X}) \geq p$ . Therefore,  $p \leq \underline{\dim}_R(\mathbf{X}) \leq \overline{\dim}_R(\mathbf{X}) \leq p$ . In other words, for this source RDD exists and is equal to  $\dim_R(\mathbf{X}) = p$ . Hence, the condition of Theorem 2 holds and we have  $\dim_R(\mathbf{X}) = \bar{d}_o(\mathbf{X})$ . This agrees with the ID of this source found in

Theorem 2 in [11],  $\bar{d}_o(\mathbf{X}) = \underline{d}_o(\mathbf{X}) = p$ . ■

Corollary 1 states that the RDD of the piecewise constant source described in Theorem 3 is equal to  $p$ , which is also the ID of this process [11]. While [11] directly computes the ID of such processes, Theorem 2 provides an easier alternate method for computing the ID as suggested in Remark 1.

## V. CONCLUSIONS

In this paper we have defined the RDD of stationary processes, as a generalization of the RDD of stochastic vectors introduced in [12]. We have proved that under some mild conditions the RDD of a stationary process is equal to its ID introduced in [11]. This gives an operational method to evaluate the ID of a stationary process, which was previously shown to be related to the fundamental limits of compressed sensing [1], [10], [11].

## REFERENCES

- [1] F. Ebrahim Rezagah, S. Jalali, E. Erkip, and H. V. Poor. Compression-based compressed sensing. *arXiv preprint arXiv:1601.01654*, 2016.
- [2] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, New York, second edition, 2006.
- [3] A. Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1-2):193–215, 1959.
- [4] D.L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.
- [5] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.
- [6] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inf. Theory*, 56(4):1982–2001, Apr. 2010.
- [7] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Found. of Comp. Math.*, 12(6):805–849, 2012.
- [8] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Process.*, 50(6):1417–1428, Jun. 2002.
- [9] D. L. Donoho, H. Kakavand, and J. Mammen. The simplest solution to an underdetermined system of linear equations. In *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pages 1924–1928, Jul. 2006.
- [10] Y. Wu and S. Verdú. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Trans. Inf. Theory*, 56(8):3721–3748, Aug. 2010.
- [11] S. Jalali and H. V. Poor. Universal compressed sensing of Markov sources. *arXiv preprint arXiv:1406.7807*, 2014.
- [12] T. Kawabata and A. Dembo. The rate-distortion dimension of sets and measures. *IEEE Trans. Inf. Theory*, 40(5):1564–1572, Sep. 1994.
- [13] T. Berger. *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [14] R. G. Gallager. *Information Theory and Reliable Communication*. NY: John Wiley, 1968.
- [15] A. D. Wyner and J. Ziv. Bounds on the rate-distortion function for stationary sources with memory. *IEEE Trans. Inf. Theory*, 17(5):508–513, 1971.
- [16] A. György, T. Linder, and K. Zeger. On the rate-distortion function of random vectors and stationary sources with mixed distributions. *IEEE Trans. Inf. Theory*, 45(6):2110–2115, Sep. 1999.
- [17] R. Gray. Rate distortion functions for finite-state finite-alphabet Markov sources. *IEEE Trans. Inf. Theory*, 17(2):127–134, Mar. 1971.
- [18] T. Berger. Explicit bounds to  $r(d)$  for a binary symmetric Markov source. *IEEE Trans. Inf. Theory*, 23(1):52–59, Jan 1977.
- [19] S. Jalali and T. Weissman. New bounds on the rate-distortion function of a binary Markov source. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 571–575. IEEE, 2007.